基于 MapReduce 模型的侵蚀地形因子计算方法研究 *

王　猛，张宏鸣†

(西北农林科技大学 信息工程学院, 陕西 杨凌 712100)

摘　要：针对传统侵蚀地形因子提取方法在处理海量数据时出现的瓶颈，提出一种基于 MapReduce 模型的侵蚀地形因子计算方法。该方法将并行计算模型 MapReduce 与改进的通用土壤流失方程（revised universal soil loss equation，RUSLE）相结合。利用最大坡降原理和 B+树建立流向关系查找树来表现地形数据的相关性；利用 MapReduce 模型进行流路查找与栅格汇聚来替代传统正反向遍历算法，解决侵蚀地形因子计算过程中汇水和累计坡长的计算效率问题。实验结果表明，对于基于海量数字高程模型数据的地形因子提取，该方法能够在计算精度允许的范围内有效提高效率。

关键词：MapReduce 模型；大数据；Hadoop；地形因子；地理信息系统

中图分类号：TP399　　doi: 10.3969/j.issn.1001-3695.2017.12.0806

Calculation of topographic factor on soil erosion based on MapReduce

Wang Meng, Zhang Hongming†

(*College of Information Engineering*, *Northwest A&F University*, *Yangling Shaanxi* 712100, *China*)

Abstract: Aiming at the bottleneck of traditional extraction method of topographic factor on soil erosion in processing massive data, we propose a method based on MapReduce. This method combined the parallel computing model MapReduce with the revised universal soil loss equation(RUSLE) . The method established the flow relationship search tree by the principle of steep slope and the B+ tree, and represented the correlation of topographic data; The method used the MapReduce model to search flow path and converge grid instead of the traditional forward and backward traversal algorithm, and solved the efficiency of the flow accumulation and slope length calculation in topographic factor calculation. The experimental results show that for the extraction of topographic factor from massive digital elevation model, the method can effectively improve the extraction efficiency in the range of the allowable accuracy.

Key words: MapReduce model; big data; Hadoop; topographic factor; geographic information system

0　引言

地形是土壤侵蚀和水土保持措施布设的主要影响因素[1]。地形特征包括坡度、坡长、坡向、曲率、地形起伏度等。在众多的地形特征中坡度和坡长是两个非常重要的特征[2]。两者影响坡面径流的汇量和水流挟沙力，是定量计算土壤流失的重要指标[3]。在通用水土流失方程和其改进版本（universal soil loss equation/revised universal soil loss equation，USLE/RUSLE）中两者共同决定了侵蚀地形因子的计算。对于小流域或者更大的区域尺度，坡度和坡长通常是利用 DEM（digital elevation model）来提取[4]。

近年，DEM 获取技术发展迅速，无人机技术、航空摄影测量、LiDar（light detection and ranging techniques）都能够方便、快捷地获得较高精度的 DEM[5]。从发展趋势来讲，DEM 数据精度不断提高，数据量不断增大，如何处理和利用这些数据已经成为地理信息系统用户的一大挑战[6]。

部分研究者认识到，现有串行算法在处理大区域海量数据时计算能力不足，因此尝试利用高性能计算平台进行水文分析。Wendleder 等人[6]利用基于消息传递接口的并行化工具，提出了一种格网 DEM 的侵蚀学坡长并行计算方法。但是该方法在选取数据分割块、设置缓冲区大小以及合并分割结果时会遇到困难[7]。Qin 等人[8]利用图形处理单元的处理能力，结合图论的并行化策略，对多流向算法进行了改进，有效提升了流向与坡度的计算效率；但受存储容量限制，该方法对海量数据地形特征提取不够理想。Dan 等人[9]借助用户级别的虚拟内存系统，利用共享内存多线程系统来处理水文地形；但该方法不利于计算像汇水这样需要反复迭代的水文特征。因此，迫切需要研究一种从海量数据中提取侵蚀地形因子的方法。

收稿日期：2017-12-13；**修回日期：**2018-02-05　**基金项目：**国家自然科学基金资助项目（41771315，41301283、41371274）；国家重点研发计划项目（2017YFC0403203）；欧盟地平线 2020 研究与创新计划项目（GA：635750）

作者简介：王猛（1992-），男，河北省承德人，硕士研究生，主要研究方向为大数据管理与分析；张宏鸣（1979-），男（通信作者），内蒙古赤峰人，副教授，主要研究方向为空间大数据管理与区域土壤侵蚀评价（zhm@nwsuaf.edu.cn）.

云计算与大数据技术的兴起为空间大数据计算带来了机遇与挑战[10]。将云计算技术与数字地形分析技术相结合，实现对海量数据的检索查询和计算，逐渐成为地理信息系统（geographic information system，GIS）的新兴研究领域[11]。目前，比较流行的开源地理空间大数据分析工具有 SpatialHadoop[12]和 GeoSpark[13]。GeoSpark 完全基于内存计算，速度上优于 SpatialHadoop。两者提供了可直接调用的接口，如范围查询、kNN（k nearest neighbor）计算和空间数据拼接等。但水文相关模型的解决方案较少，尤其对于侵蚀因子计算过程中汇水及累积坡长的提取还没有提供解决方案。

本文利用 MapReduce 模型的分布式计算能力，提出一种分布式侵蚀地形因子计算方法。该方法解决了海量地形数据计算过程中出现的效率低下和内存限制问题，对于促进数字地形分析理论与技术体系的完善和 SpatialHadoop 和 GeoSpark 等工具的发展有借鉴意义。

1　侵蚀地形因子计算原理

侵蚀地形因子提取过程主要分为四个部分：a）利用 Deterministic 8（D8）算法计算流向和坡度；b）依据 a）中流向数据计算汇水面积，并利用阈值和截断因子及流向数据设置流路中断点；c）计算累计坡长；d）根据地形学公式，利用 a）中坡度和 c）中累计坡长计算侵蚀地形因子。其中输入的 DEM 数据是 ASCII 文本数据。为了保证数据的完整性，在计算之前需要判断数据是否符合计算要求。

1.1　D8 算法

坡度和流向的计算采用 O'Callaghan 等人[14]提出的 D8 算来计算。D8 算法的原理是：用最大坡降原理来确定水流的方向，即在 3*3 的 DEM 网格上，计算中心网格与各相邻网格间的坡度，取坡度结果最大值作为中心网格的坡度，并且最大值所在方向即为中心网格的流向[15]。坡度计算公式如式（1）所示。其中：*S* 为坡度；DEM_c 为中心栅格的高程值；DEM_i 为周围 8 个栅格的高程值；*Length* 在水平或竖直方向取 *CellSize*，在对角线方向取 $\sqrt{2}$ **CellSize*。

$$S=\begin{cases}Max\left(\arctan\left(\dfrac{DEM_c-DEM_i}{Length}\right)\right) & DEM_c\neq DEM_i\\ 0.1 & DEM_c=DEM_i\end{cases}\tag{1}$$

1.2　汇水计算及截断设置

汇水的计算是根据每个栅格的水流方向，计算出每一个栅格单元的上游集水面积[16]。目前比较常用的是 Zhang 等人[16]提出的迭代算法。该方法采用扫描线的方式，通过对栅格点正反向遍历计算汇水。

坡长计算需要考虑两种截断：坡度截断和沟道截断。坡度截断通过判断坡度变化率和中断因子的关系来确定。当坡度变化率大于中断因子时，该点标记为截断点。沟道截断通过设置阈值的方法确定，当汇水面积大于阈值时，标记该点为截断点。

1.3　累计坡长计算

计算累计坡长之前需要计算每个栅格的 USL（unit slope length）。USL 主要有以下两种情形：a）基础栅格，如果栅格的流向是水平或垂直方向，取值 *CellSize*，如果是对角线方向，取值 $\sqrt{2}$ **CellSize*；b）截断栅格，如果该截断栅格的流向是水平或垂直方向，取值 *CellSize/2* ，如果是是对角线方向，取值 $\sqrt{2}$ /2**CellSize*。

累计坡长为从坡面径流的起点到径流被拦截点或流路中断点的水平距离[17]。其计算公式为如式（2）所示。其中：$\lambda_{i,j}$ 为坐标点（*i*，*j*）的坡长；*m* 是坡长指数。

$$\lambda_{i,j}=\sum_{x=0,y=0}^{x=i,y=j}\sum_{k=1}^{m}\lambda_{x,y}\tag{2}$$

1.4　侵蚀地形因子计算

本文利用 Mccool 等人[18]提出的式(3)~(7)完成侵蚀地形因子的计算。其中：*LS* 表示侵蚀地形因子；*S* 表示坡度因子；*L* 表示坡长因子；θ 代表坡度；λ 代表坡长；22.1 表示标准小区坡长；β 表示坡度变化因子。

$$LS=L*S\tag{3}$$

$$S=\begin{cases}10.8\sin\theta+0.03 & \theta<9^\circ\\16.8\sin\theta-0.05 & \theta\geq 9^\circ\end{cases}\tag{4}$$

$$L=(\lambda/22.1)^m\tag{5}$$

$$m=\beta/(1+\beta)\tag{6}$$

$$\beta=(\sin\theta)/\left[3\cdot(\sin\theta)^{0.8}+0.56\right]\tag{7}$$

2　基于 MapReduce 模型的侵蚀地形因子计算

2.1　实验数据

为测试实验结果的准确性，本文利用陕西省韭园沟和甘肃省龙泉县不同分辨率的 DEM 作为研究数据进行测试。韭园沟面积约 70 km^2，海拔 825 m~1200 m，平均坡度 28.9°。龙泉县面积约 100 km^2，海拔 1951 m~2540 m，平均坡度 26.6°。两块区域属于典型的黄土丘陵沟壑区，地表支离破碎，水土流失严重，具有较大的高程范围和坡度范围，适合做常规测试。此外，为验证程序在大区域尺度的运行情况，本文使用全国 SRTM（shuttle radar topography mission）1 弧秒（分辨率约 30 m）数据进行测试（数据来源：https://e4ftl01.cr.usgs.gov/）。SRTM 是美国太空总署（NASA）和国防部国家测绘局（NIMA）以及德国与意大利航天机构共同合作完成联合完成地球表面所进行的遥感测量。这是全球数字测绘的一个重大突破，为全球范围内的地形分析提供了高质量的 DEM 数据[19]。全国 SRTM 如图 1 所示。

2.2　方法总体设计

根据侵蚀地形因子的计算特征，MapReduce 模型下侵蚀地形因子计算可以分为三个部分（图 2（b））：a）利用 D8 算法和 B+树建立流向关系查找树，并将地形数据存储为适合 MapReduce 输入的 key-vaule 形式;b）利用 MapReduce 模型分

布式计算汇水及累计坡长；c）用地形学公式计算侵蚀地形因子，并将结果以 DEM 形式存储。

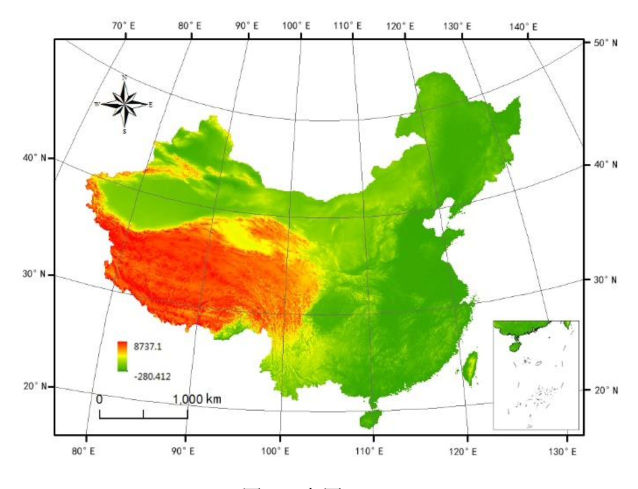
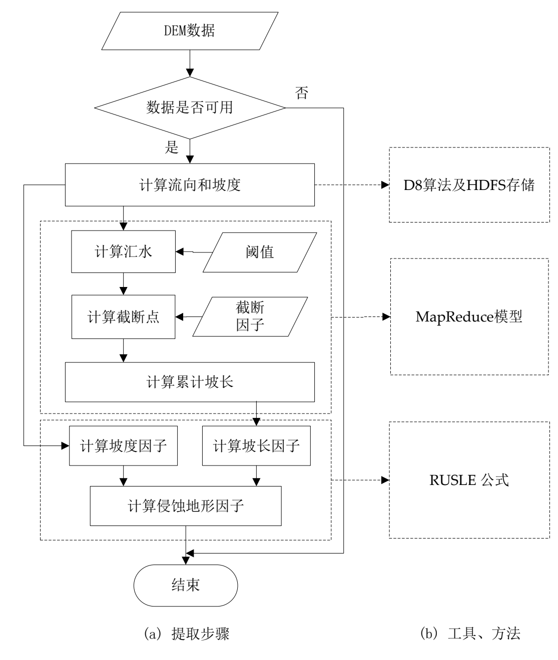


图 1　全国 SRTM



(a) 提取步骤　　(b) 工具、方法

图 2　侵蚀地形因子计算过程

2.3　HDFS 存储

MapReduce 的输入利用 FileInputFormat 组件完成，该组件从 HDFS（hadoop distributed file system）上读取文件，并向 MapReduce 输出的以行号为 key，以该行内容为 value 的键值对[20]。DEM 数据是用一组有序数值阵列形式表示地面高程的一种实体地面模型[16]。由于 DEM 数据中各栅格点之间的相关性，并不适合直接用于 MapReduce 计算。为此需要建立流向关系映射表和将地形数据以行级形式存储在 HDFS 上。其计算过程如图 3 所示。

1）利用 B+树建立流向关系查找树。

本文利用 1.1 节中的 D8 算法对 DEM 数据进行计算，并利用 B+树建立流向关系查找树，这样 DEM 网格中的每一点都可以快速获取其流出方向。该查找树具有如下特征：a)所有节点都是以栅格点的坐标位置为关键字，并且叶子节点包含关键字所在位置栅格的流出信息；b)在插入过程中，当节点中关键字个数大于 1 000 时分裂成两个节点；c)查找是通过判断与关键字关系，选择相应路径，直到找到叶子节点获取流出信息。

2）地形数据以行级形式存储到 HDFS。

在计算流向坡度过程中，将地形中每一个栅格点以字符串“<坐标信息> <高程值> <坡度值> <栅格尺寸> <截断信息> <汇水值><坡长><侵蚀地形因子>”形式存储为文件的一行。然后利用 HDFS 客户端将文件存放到 HDFS 上，HDFS 的 NameNode 节点自动将文件分割为 64 MB 大小的 Block，建立冗余副本，并且数据块存储在一组 DataNode 节点中[21]。

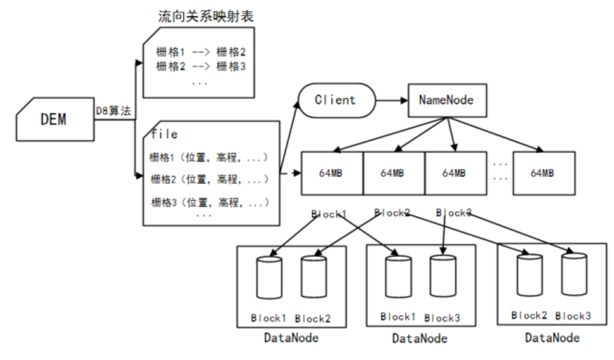


图 3　DEM 数据转换过程

2.4　MapReduce 模型计算汇水和累计坡长

MapReduce 是一个用于处理大规模数据集的并行计算模型[22]。该模型由 Map 和 Reduce 两个阶段组成。每个阶段都是用键值对（key/value）作为输入（input）和输出（output）[23]。Map 运行在多个计算节点上，对 HDFS 上一个或多个数据块处理；Map 的输出在模型内部会经历 Shuffle 过程，即排序、汇总，会把相同 key 值的记录统计到同一起；Reduce 是对 Shuffle 后的数据合并输出。而程用户要做的就是定义好 map 函数和 reduce 函数，即可完成分布式计算[24]。

传统汇水和累计坡长计算利用正反向遍历方法完成，是程序中最为耗时的部分。利用 MapReduce 模型进行流路查找与栅格汇聚，来解决汇水和累计坡长的计算效率问题。

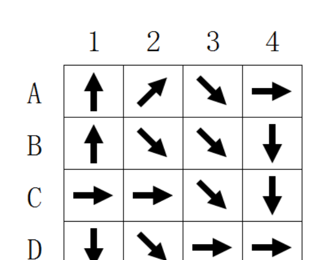
2.4.1 计算汇水

汇水计算的 map 函数与 reduce 函数的伪代码如下：

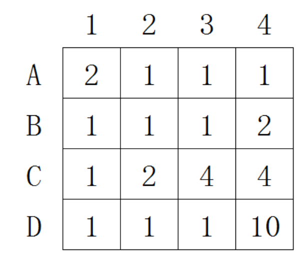
```
map(LongWritable key, Text value){
    栅格对象 O = new 栅格对象(value）;
    For each 栅格 T which O flow into{
        Collect(T,栅格尺寸);
    }
}
reduce(栅格对象 K, LongWrite []v){
    long count = 0;
    For each v in V{
        count += v;
    }
    Collect(K,count);
```

```
}
```

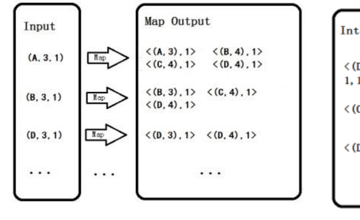
Map 计算的输入 2.3 节存储在 HDFS 的地形数据，其输入是以行号为 key，所在行字符串为 value 的键值对。其计算过程是将字符串转换为栅格对象点，并对于每个栅格点，利用流向查找树去查询以该点为起点的流路，并将查找结果以 key-value 形式输出，其中 key 为当前查找到的栅格对象，值为栅格尺寸。其计算过程如图 4 所示（设栅格尺寸为 1），图 4（a）表示每个点的流向；（c）是 Map 计算过程，如当前计算节点为 B3，根据流向关系知，B3 流向 C4，C4 流向 D4，D4 流向无值区，因此，输出<(B,3),1>、<(C,4),1>和<(D,4),1>这样三个键值对。其他点计算过程与该点相同。



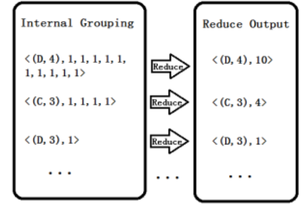
(a)流向数据



(b)汇水结果



(c)Map 过程



(d)Reduce 过程

图 4　汇水计算过程

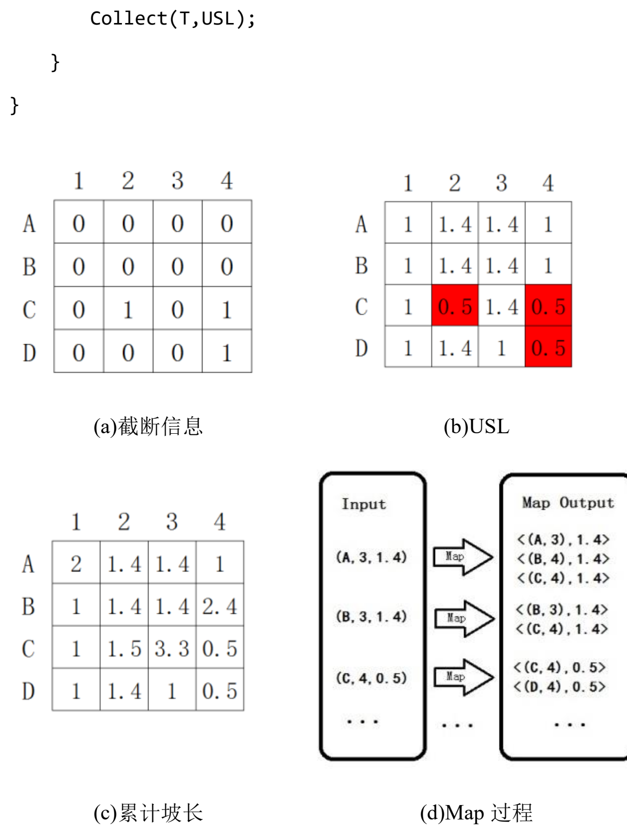
Map 计算结束后，Shuffle 过程进行排序、汇总，把 key 值相同的 value 进行统计，输出到 Reduce。汇水的 Reduce 过程就是把同一个 key 值下的各个 value 值累计起来，累计和即为 key 所在位置的汇水面积。Reduce 的输出 key 为栅格对象点，值为汇水面积的键值对。其计算过程如图 4（d）所示，对于 C3 点，其 value 值是一个包含 4 个 1 的集合，集合值累计起来得到 4，该值就是 C3 点的汇水。所有 Reduce 过程结束后的汇水结果如图 4（b）所示。

2.4.2 计算累计坡长

计算累计坡长之前需设置截断信息和计算 USL。如图 5（b）是根据图 5（a）的截断信息（其中 1 表示截断）和图 4（a）的流向信息计算出来的 USL。

累计坡长的 MapReduce 计算过程与汇水计算类似，但稍有不同。其中 Map 计算时需要考虑截断信息；而 Reduce 计算与汇水的 Reduce 计算原理完全相同。map 函数的伪代码如下：

```
map(LongWritable key, Text value){
    栅格对象 O = new 栅格对象(value）;
    For each 非截断栅格 T which O flow into{
        Collect(T,USL);
    }
}
```



(a)截断信息　　(b)USL

(c)累计坡长　　(d)Map 过程

图 5　累计坡长计算过程

累计坡长的 Map 过程输入是以行号为 key，所在行字符串为 value 的键值对。其计算过程是将字符串转换为栅格对象点，并每个栅格点，利用流向查找树去查询以该点为起点到截断点的流路，并将查找结果以 key-value 形式输出，其中 key 为当前查找点，value 为流路起点的 USL。其计算过程如图 5（d）所示。假设当前计算点为 A3 点，USL 为 1.4，根据流向关系，A3 流向 B4，B4 流向 C4，C4 是截断点，不再寻找，因此，输出<(A,3),1.4>、<(B,4),1.4>和<(C,4),1.4>这样三个键值对。需要注意一种特殊情况，当输入点本身为截断点时，需要继续寻找流路。如当前计算点为截断点 C2，USL 为 0.5，根据流向关系，C2 流向 C4，C4 为截断点，不再寻找，因此输出键值对<(C,2),0.5>和<(C,4),0.5>。

累计坡长的 Reduce 过程与汇水的 Reduce 过程计算方式相同，都是对同一个 key 值下 value 值累计，累积和为累计坡长。Reduce 的输出 key 为栅格对象点，值为累计坡长的键值对。图 5（c）为最终累计坡长计算结果。

2.5　侵蚀地形因子计算

地形中每个栅格点的坡度和累计坡长都以行级形式存储在 HDFS 上。读取文件每一行，转换为栅格对象，利用式(3)~(7)完成侵蚀地形因子的计算。

2.6　实验平台

本文利用 Hadoop 平台实现侵蚀地形因子计算。Hadoop 是一个开发和处理海量数据的软件平台，能够利用多台计算机组成的集群对海量数据进行分布式计算[25]。实验利用 6 台计算机搭建集群，每台计算机为 64 位 Ubuntu 操作系统、64 GB 内存、

4 核 Intel (R) Core (TM) i5 CPU 处理器。

2.7　精度对比

为验证本文方法的准确性，实验使用标准差（standard deviation，SD）和绝对差（absolute deviation，AD）比较了 DWSEL（distributed watershed erosion slope length）方法[16]与本文方法在坡度、坡长和侵蚀地形因子的差异。其计算公式如式（8）和（9）所示。

$$SD=\sqrt{\frac{1}{N}\sum_{1}^{N}\left(R_{this\ paper}-R_{DWSEL}\right)^{2}} \tag{8}$$

$$AD=\frac{1}{N}\sum_{1}^{N}\left|R_{this\ paper}-R_{DWSEL}\right| \tag{9}$$

其中：N 为 DEM 数据的栅格数；R 是比较类型，可取坡度、坡长以及侵蚀地形因子值。

3　实验结果与分析

3.1　精度分析

本文方法与 DWSEL 方法在坡度、坡长和侵蚀地形因子的差异如表 1 所示。可以看到：a) 两种方法的坡度计算结果完全相同；b)两种方法的坡长结果存在一定差异，统计发现坡长误差在 10 m 以内，根据 McCool[26]等的规定，该误差在允许范围内；c) 侵蚀地形因子由坡度和坡长共同决定，结果非常接近。

表 1　本文方法与 DWSEL 方法在坡度、坡长及侵蚀地形因子之间的差异

区域	分辨率（m）	坡度		坡长		侵蚀地形因子	
		标准差	绝对差	标准差	绝对差	标准差	绝对差
韭园沟	1	0.000	0.000	0.027	0.026	0.001	0.001
	2	0.000	0.000	0.056	0.035	0.003	0.002
龙泉	1	0.000	0.000	0.040	0.033	0.002	0.002
	2	0.000	0.000	0.074	0.040	0.006	0.005
全国 SRTM	30	--[1]	--[1]	--[1]	--[1]	--[1]	--[1]

注：[1] 表示 DWSEL 方法无法计算。

3.2　效率分析

本文方法与 DWSEL 方法在计算流向和坡度、计算汇水面积和累计坡长、计算侵蚀地形因子的时间以及总计算时间对比如表 2 所示。可以看到：a) 对于韭园沟和龙泉两块区域，本文方法平均需要花费 DWSEL 方法的 85.7%时间处理流向和坡度，90.0%的时间计算侵蚀地形因子，这两部分效率相差不大，在计算汇水和累计坡长步骤中，本文方法平均只需要花费 DWSEL 方法的 11.7%的时间；b) 对于全国 SRTM 数据（单个文件达 150 GB），DWSEL 无法计算完成，本文方法通过 MapReduce 模型，可以在多个计算节点上分布式计算完成，提升了计算效率。

3.3　全国 SRTM 分析

全国 SRTM 数据的侵蚀地形因子结果如图 6 所示。对于全国 SRTM 数据，DWSEL 方法无法计算。为此对全国主要水蚀区的侵蚀地形因子进行统计。北方土石山区侵蚀地形因子值主要在 0.5~20 间。东北漫岗丘陵区侵蚀地形因子值在 0.5 以下占 76.5%，没有因子值为 5 以上的；西南紫色土丘陵区侵蚀地形因子值主要在 5 以上，0.5 以下的占 5.1%；西北黄土丘陵区侵蚀地形因子值主要在 5 以上，0.5 以下的占 1.3%；南方红壤丘陵区侵蚀地形因子值在 0.5 以下占 17%，因子值大于 20 占 21%左右；与杨勤科等[19,27]对全国典型水蚀区地形分析结果基本一致。

表 2　本文方法与 DWSEL 方法运行时间

区域	分辨率(m)	栅格数（行*列）	方法	运行时间 (s)			
				流向和坡度	汇水和累计坡长	侵蚀地形因子	总时间
韭园沟	1	11110*12550	DWSEL	813.371	4039.338	222.047	5354.224
			本文方法	616.191	331.676	173.166	1420.331
	2	5555*6275	DWSEL	182.199	725.522	48.444	1147.398
			本文方法	167.636	68.415	47.033	323.089
龙泉	1	15000*12000	DWSEL	709.856	3087.986	199.234	4198.367
			本文方法	586.065	394.089	155.651	1285.805
	2	7500*6000	DWSEL	154.531	519.044	43.221	874.574
			本文方法	143.085	84.162	46.112	299.369
全国 SRTM	30	216439*137962	DWSEL	--1	--1	--1	--1
			本文方法	29345.270	11471.105	4008.356	48606.003

[1] DWSEL 方法无法计算

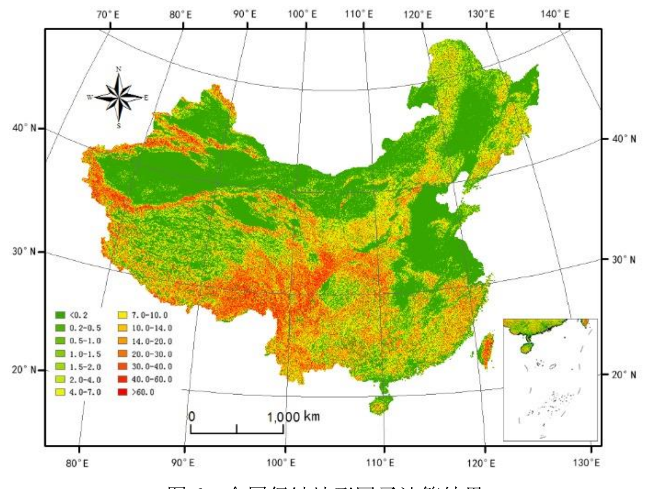


图 6　全国侵蚀地形因子计算结果

4　结束语

本文利用 MapReduce 模型的分布式计算能力，提出一种分布式侵蚀地形因子计算方法。该方法利用最陡坡原理和 B+树建立流向关系查找树，利用 MapReduce 模型计算汇水和累计坡长来解决侵蚀地形因子计算的效率问题和内存限制。实验结果表明，对于基于海量数字高程模型数据的地形因子提取，该方法能够在计算精度允许的范围内有效提高效率。但该方法利用的是单流向算法，如何利用多流向算法更加精确地提取侵蚀地形因子是今后的研究问题之一。

参考文献：

[1] 张宏鸣，宋泽鲁，杨江涛，等. DEM 超分辨率重构对梯田坡度提取的影响研究 [J]. 农业机械学报, 2017 (1): 112-118.

[2] Liu K, Tang G, Jiang L, *et al.* Regional-scale calculation of the LS factor using parallel processing [J]. Computers & Geosciences, 2015, 78 (C): 110-122.

[3] Kong Y, 张科利, Cao L. 土壤侵蚀研究中的坡长因子评价问题 [J]. 水土保持研究, 2008, 15 (4): 43-47.

[4] 杨勤科，李锐，曹明明. 区域土壤侵蚀定量研究的国内外进展 [J]. 地球科学进展, 2006, 21 (8): 849-856.

[5] Bai R, Li T, Huang Y, *et al.* An efficient and comprehensive method for drainage network extraction from DEM with billions of pixels using a size-balanced binary search tree [J]. Geomorphology, 2015, 238: 56-67.

[6] Wendleder A, Felbier A, Wessel B, *et al.* A method to estimate long-wave height errors of SRTM C-band DEM [J]. IEEE Geoscience & Remote Sensing Letters, 2016, 13 (5): 696-700.

[7] Byun J, Seong Y B. An algorithm to extract more accurate stream longitudinal profiles from unfilled DEMs [J]. Geomorphology, 2015, 242: 38-48.

[8] Qin C, Zhan L. Parallelizing flow-accumulation calculations on graphics processing units—from iterative DEM preprocessing algorithm to recursive multiple-flow-direction algorithm [J]. Computers & Geosciences, 2012, 43 (6): 7-16.

[9] Dan W, Tarboton D, Wallace R M. A virtual tile approach to raster-based calculations of large digital elevation models in a shared-memory system [J]. Computers & Geosciences, 2015, 82 (C): 78-88.

[10] Zhao L, Chen L, Ranjan R, *et al.* Geographical information system parallelization for spatial big data processing: a review [J]. Cluster Computing, 2016, 19 (1): 139-152.

[11] 温馨，罗侃，陈荣国. 基于 Shark//Spark 的分布式空间数据分析框架 [J]. 地球信息科学学报, 2015, 17 (4): 401-407.

[12] Eldawy A, Mokbel M F. SpatialHadoop: a MapReduce framework for spatial data [C]// Proc of IEEE, International Conference on Data Engineering. 2016: 1352-1363.

[13] Yu J, Wu J, Sarwat M. GeoSpark: a cluster computing framework for processing large-scale spatial data [C]// Proc of Sigspatial International Conference on Advances in Geographic Information Systems. 2015: 70.

[14] O'Callaghan J F, Mark D M. The extraction of drainage networks from digital elevation data [J]. Computer Vision Graphics & Image Processing, 1984, 28 (3): 323-344.

[15] 孙崇亮，王卷乐. 基于 DEM 的水系自动提取与分级研究进展 [J]. 地理科学进展, 2008, 27 (1): 118-124.

[16] Zhang H, Yang Q, Li R, *et al.* Extension of a GIS procedure for calculating the RUSLE equation LS factor [J]. Computers & Geosciences, 2013, 52 (1): 177-188.

[17] 张宏鸣，杨勤科，李锐，等. 流域分布式侵蚀学坡长的估算方法研究 [J]. 水利学报, 2012, 43 (4): 437-444.

[18] Mccool D K, Brown L C, Foster G R, *et al.* Revised slope length factor for the universal soil loss equation [J]. Transactions of the Asae, 1989, 30 (5): 1387-1396.

[19] 郭明航，杨勤科，王春梅. 中国主要水蚀典型区侵蚀地形特征分析 [J]. 农业工程学报, 2013, 29 (13): 81-89.

[20] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters [C]// Proc of Conference on Symposium on Opearting Systems Design & Implementation. USENIX Association, 2004: 10-10.

[21] 廖彬，于炯，张陶，等. 基于分布式文件系统 HDFS 的节能算法 [J]. 计算机学报, 2013 (5): 1047-1064.

[22] 宋杰，孙宗哲，毛克明，等. MapReduce 大数据处理平台与算法研究进展 [J]. 软件学报, 2017 (3): 514-543.

[23] 肖文，胡娟，周晓峰. 基于 MapReduce 计算模型的并行关联规则挖掘算法研究综述 [J]. 计算机应用研究, 2018, 35 (1): 13-23.

[24] 宋杰，徐澍，郭朝鹏，等. 一种优化 MapReduce 系统能耗的任务分发算法 [J]. 计算机学报, 2016 (2): 323-338.

[25] 刘磊，尹芳，冯敏，等. 基于开源 Hadoop 的栅格数据分布式处理 [J]. 华中科技大学学报：自然科学版, 2013, 41 (7): 103-108.

[26] Mccool D K, Foster G R, Weesies G A. Slope length and steepness factors (LS) [J]. 1997.

[27] 杨勤科，郭明航，李智广，等. 全国土壤侵蚀地形因子提取与初步分析 [J]. 中国水土保持, 2013 (10): 17-2.